

```

1 AACAGCACGA GGGCGAGGGG ACGTCTCCTC TCCCCCAGCT GCTCTGCTCG
51 GATGGCGCCG CCGGCTGAGT GACGGGGGCG GCGCGCAGGA CTTCCCAGCT
101 CGGACCTCTT GCCTTCGAGG GGAAAGATGT ACGAGAGTGT AGAAGTGGGG
151 GGTCACCCC CTAATCCCTT CCTAGTGGTG GATTTTATA ACCAGAACCG
201 GGCCGTGTTG CTCCAGAGA AGGGGCTCCC CGCCCCGGGT CCGTACTCCA
251 CCCCCTCCG GACTCCGCTT TGGAATGGCT CAAACCACTC CATTGAGACC
301 CAGAGCAGCA GTTCTGAAGA GATAGTGCCC AGCCCTCCCT CGCCACCCCC
351 TCTACCCCGC ATCTACAAGC CTTGCTTTGT CTGTCAGGAC AAGTCCTCAG
401 GCTACCACTA TGGGGTCAGC GCCTGTGAGG GCTGCAAGGG CTTCTTCCGC
451 CCGCAGCATC AGAAGAACAT GGTGTACACG TGTCACCGGG ACAAGAACTG
501 CATCATCAAC AAGGTGACCC GGAACCCCTG CCAGTACTGC CGACTGCAGA
551 AGTGCTTTGA AGTGGGCATG TCCAAGGAGT CTGTGAGAAA CGACCGAAAC
601 AAGAAGAAGA AGGAGGTGCC CAAGCCCGAG TGCTCTGAGA GCTACACGCT
651 GACGCCGGAG GTGGGGGAGC TCATTGAGAA GGTGCGCAA GCGCACCAGG
701 AAACCTTCCC TGCCCTCTGC CAGCTGGGCA AATACACTAC GAACAACAGC
751 TCAGAACAAAC GTGTCTCTCT GGACATTGAC CTCTGGGACA AGTTCAGTGA
801 ACTCTCCACC AAGTGCATCA TTAAGACTGT GGAGTTCGCC AAGCAGCTGC
851 CCGGCTTCAC CACCCTCACC ATCGCCGACC AGATCACCTT CCTCAAGGCT
901 GCCTGCCTGG ACATCCTGAT CCTGCGGATC TGACGCGGT ACACGCCCGA
951 GCAGGACACC ATGACCTTCT CGGACGGGCT GACCCTGAAC CGGACCCAGA
1001 TGCACAACGC TGGCTTCGGC CCCCTCACC ACCTGGTCTT TGCCCTTCGCC
1051 AACCAGCTGC TGCCCTGGA GATGGATGAT GCGGAGACGG GGCTGCTCAG
1101 CGCCATCTGC CTCATCTGCG GAGACCGCCA GGACCTGGAG CAGCCGGACC
1151 GGGTGACAT GCTGCAGGAG CCGCTGCTGG AGGCGCTAAA GGTCTACGTG
1201 CGGAAGCGGA GGCCAGCCG CCCCACATG TTCCCCAAGA TGCTAATGAA
1251 GATTACTGAC CTGCGAAGCA TCAGCGCCA GGGGGCTGAG CGGGTGATCA
1301 CGCTGAAGAT GGAGATCCCG GGCTCCATGC CGCTCTCAT CCAGGAAATG
1351 TTGGAGAACT CAGAGGGCCT GGACACTCTG AGCGGACAGC CGGGGGGTGG
1401 GGGGCGGGAC GGGGGTGGCC TGCCCCCCCC GCCAGGCAGC TGTAGCCCCA
1451 GCCTCAGCCC CAGCTCCAAC AGAAGCAGCC CGGCCACCCA CTCCCCGTGA
1501 CCGCCACGC CACATGGACA CAGCCCTCGC CCTCCGCCCC GGCTTTTCTC
1551 TGCCTTTCTA CCGACCATGT GACCCCGCAC CAGCCCTGCC CCCACCTGCC
1601 CTCCCGGGCA GTACTGGGGA CCTTCCCTGG GGGACGGGGA GGGAGGAGGC
1651 AGCGACTCCT TGGACAGAGG CCTGGGCCCT CAGTGGACTG CCTGCTCCCA
1701 CAGCCTGGGC TGACGTCAGA GGCCGAGGCC AGGAACTGAG TGAGGCCCTT
1751 GGTCTGGGT CTCAGGATGG GTCTGGGGG CCTCGTGTTT ATCAAGACAC
1801 CCCTCTGCCC AGCTCACCAC ATCTTCATCA CCAGCAAACG CCAGGACTTG
1851 GCTCCCCCAT CCTCAGAACT CACAAGCCAT TGCTCCCCAG CTGGGGAACC
1901 TCAACCTCCC CCCTGCCTCG GTTGGTGACA GAGGGGGTGG GACAGGGGCG
1951 GGGGGTTCCT CCTGTACATA CCCTGCCATA CCAACCCAG GTATTAATTC
2001 TCGCTGGTTT TGTTTTTATT TTAATTTTTT TGTTTTGATT TTTTAAATAA
2051 GAATTTTCAT TTTAAGCACA AAAAAAAAAA AAAAAA

```

FEATURES:

Start codon: 127

Stop codon: 1498

Homologous proteins:

Top 10 BLAST Hits

	Score	E
gi 3213188 gb AAC23439.1 (U15211) retinoic acid receptor alpha...	918	0.0
gi 4506419 ref NP_000955.1 retinoic acid receptor, alpha >gi 1...	830	0.0
gi 35874 emb CAA29787.1 (X06538) retinoic acid receptor (AA 1 ...	830	0.0
gi 7638251 gb AAF65452.1 (AF242867) ECFP-retinoic acid recepto...	830	0.0
gi 7441783 pir S78481 retinoic acid receptor alpha-2 - eastern...	829	0.0
gi 4160009 gb AAD05222.1 (AF088895) retinoic acid receptor alp...	829	0.0
gi 1314308 gb AAB00112.1 (U41742) nucleophosmin-retinoic acid ...	827	0.0
gi 8815561 gb AAB19602.2 (S50916) retinoic acid receptor alpha...	827	0.0
gi 1314310 gb AAB00113.1 (U41743) nucleophosmin-retinoic acid ...	827	0.0
gi 190125 gb AAA60126.1 (M73779) PML-RAR protein [Homo sapiens]	827	0.0
gi 545176 gb AAB29813.1 retinoic acid receptor alpha, RAR alph...	824	0.0
gi 2119682 pir I50674 retinoic acid receptor alpha isoform 2-1...	819	0.0
gi 1743345 emb CAA71177.1 (Y10094) retinoic acid receptor-alph...	816	0.0
gi 133484 sp P11416 RRA_MOUSE RETINOIC ACID RECEPTOR ALPHA (RAR...	815	0.0

EST:

gi 847367 gb R73335.1 R73335 yll0b08.rl Soares breast 2NbHBst H...	708	0.0
gi 2714987 gb AA705069.1 AA705069 zj83e11.sl Soares_fetal_liver...	517	e-144
gi 8636851 gb BE174125.1 BE174125 QV1-HT0572-200300-117-cl2 HT0...	466	e-128
gi 2841314 gb AA781983.1 AA781983 ai78g12.sl Soares_testis_NHT ...	297	5e-78
gi 611497 gb T29399.1 T29399 EST79267 Human Placenta Homo sapie...	123	2e-25
gi 1969924 gb AA317545.1 AA317545 EST19547 Retina II Homo sapie...	80	2e-12

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

gi 847367 gb R73335.1	Human Breast
gi 2714987 gb AA705069.1	Human fetal liver-spleen
gi 8636851 gb BE174125.1	Human head-neck
gi 2841314 gb AA781983.1	Human testis
gi 611497 gb T29399.1	Human placenta
gi 1969924 gb AA317545.1	Human retina

Expression information from PCR-based tissue screening panels

Human Brain
Human Placenta
Human Liver
Human Hela cells
Human Kidney

```

1  MYESVEVGGP TPNPFLVVDF YNQNRACLLP EKGLPAPGPY STPLRTPLWN
51  GSNHSIETQS SSSEEIVPSP PSPPLPRIY KPCFVCQDKS SGYHYGVSAC
101 EGCKGFFRRS IQKNMVTCH RDKNCIINKV TRNPCQYCRL QKCFEVGMSK
151 ESVRNDRNKK KKEVPKPECS ESYTLTPEVG ELIEKVRKAH QETFPALCQL
201 GKYTTNNSSE QRVSLDIDLW DKFSELSTKC IKTVEFAKQ LPGFTTLTIA
251 DQITLLKAAC LDILILRICT RYTPEQDTMT FSDGLTLNRT QMHNAGFGPL
301 TDLVFAFANQ LLPLEMDDAE TGLLSAICLI CGDRQDLEQP DRVDMLEPEL
351 LEALKVYVRK RRPSRPHMFP KMLMKITDLR SISAKGAERV ITLKMEIPGS
401 MPPLIQEMLE NSEGLDTLSG QPGGGGRDGG GLPPPPGSCS PSLSPSSNRS
451 SPATHSP

```

FEATURES:

Functional domains and key regions:

<u>InterPro</u>	<u>Results of FPrintScan against PRINTS</u>	<u>Results of HMMPfam against PFAM-A</u>	<u>Results of PPsearch against PROSITE</u>	<u>Results of ProfileScan against PROSITE profiles</u>
<u>IPR000003</u> Retinoic acid receptor	<u>PR00545</u> [149-163] [340-357] [361-380]			
<u>IPR000324</u> Vitamin D receptor	<u>PR00350</u> [83-99] [100-119]			
<u>IPR000536</u> Ligand-binding domain of nuclear hormone receptor		<u>PF00104</u> [225-383]		
<u>IPR001628</u> C4-type steroid receptor zinc finger	<u>PR00047</u> [83-99] [99-114] [132-140] [140-148]	<u>PF00105</u> [81-156]	<u>PS00031</u> [83-109]	
<u>IPR001723</u> Steroid hormone receptor	<u>PR00398</u> [144-154] [226-247] [247-263] [314-329] [371-388]			

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	292	312	0.641	Putative

BLAST Alignment to Top Hit:

>gi|3213188|gb|AAC23439.1| (U15211) retinoic acid receptor alpha 2
isoform [Rattus norvegicus]
Length = 459

Score = 918 bits (2347), Expect = 0.0
Identities = 445/457 (97%), Positives = 449/457 (97%), Gaps = 2/457 (0%)
Frame = +1

Query: 1 MYESVEVGG--PTPNPFLVDFYNQNRACLLPEKGLPAPGPYSTPLRTPPLWNGSNHSIET 174
MYESVEVGG P PNPFLVDFYNQNRACLL EKGLPAPGPYSTPLRTPPLWNGSNHSIET
Sbjct: 1 MYESVEVGGLTPAPNPFLVDFYNQNRACLLQEKGLPAPGPYSTPLRTPPLWNGSNHSIET 60

Query: 175 QSSSSEEIVSPSPSPPLPRIYKPCFVCQDKSSGYHYGVSACEGCKGFFRRSIQKNMVYT 354
QSSSSEEIVSPSPSPPLPRIYKPCFVCQDKSSGYHYGVSACEGCKGFFRRSIQKNMVYT
Sbjct: 61 QSSSSEEIVSPSPSPPLPRIYKPCFVCQDKSSGYHYGVSACEGCKGFFRRSIQKNMVYT 120

Query: 355 CHRDKNCIINKVTRNPCQYCRLOKCFEVGMSKESVRNDRNKKKKEVPKPECSESYTLTPE 534
CHRDKNCIINKVTRN CQYCRLOKCFEVGMSKESVRNDRNKKKKE PKPECSESYTLTPE
Sbjct: 121 CHRDKNCIINKVTRNRCQYCRLOKCFEVGMSKESVRNDRNKKKKEPKPECSESYTLTPE 180

Query: 535 VGELIEKVRKAHQETFPALCQLGKYTTNNSSEQRVSLDIDLWDKFSELSTKCIIKTVEFA 714
VGELIEKVRKA+QETFPALCQLGKYTTNNSSEQRVSLDIDLWDKFSELSTKCIIKTVEFA
Sbjct: 181 VGELIEKVRKANQETFPALCQLGKYTTNNSSEQRVSLDIDLWDKFSELSTKCIIKTVEFA 240

Query: 715 KQLPGFTTLTIADQITLLKAACLDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFG 894
KQLPGFTTLTIADQITLLKAACLDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFG
Sbjct: 241 KQLPGFTTLTIADQITLLKAACLDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFG 300

Query: 895 PLTDLVFAFANQLLPLEMDDAETGLLSAICLICGDRQDLEQPDVDMLEPLLEALKVYV 1074
PLTDLVFAFANQLLPLEMDDAETGLLSAICLICGDRQDLEQPD+VDMLEPLLEALKVYV
Sbjct: 301 PLTDLVFAFANQLLPLEMDDAETGLLSAICLICGDRQDLEQPDKVDMLQEPLLEALKVYV 360

Query: 1075 RKRRPSRPHMFPKMLMKITDLRSISAKGAERVITLKMEIPGSMPLIQEMLENSEGLDTL 1254
RKRRPS+PHMFPKMLMKITDLRSISAKGAERVITLKMEIPGSMPLIQEMLENSEGLDTL
Sbjct: 361 RKRRPSQPHMFPKMLMKITDLRSISAKGAERVITLKMEIPGSMPLIQEMLENSEGLDTL 420

Query: 1255 SGQPGGGGRDGGGLPPPPGSCSPSLSPSSNRSSPATHSP 1371
SGQ GGG RDGGGL PPPGSCSPSLSPSS+RSSPAT SP
Sbjct: 421 SGQSGGGTRDGGGLAPPPGSCSPSLSPSSHRSSPATQSP 459

006T0T"022T6960

Hammer results:

Model	Description	Score	E-value	N
CE00550	CE00550 retinoic_acid_receptor	1204.3	0	1
CE00342	E00342 retinoid_X_receptor	341.6	6.8e-99	2
PF00104	Ligand-binding domain of nuclear hormone rec	241.4	1.3e-68	1
PF00105	Zinc finger, C4 type (two domains)	192.6	1.2e-56	2
CE00055	CE00055 peroxisome_proliferator_activated_re	132.1	4.9e-45	3
CE00059	CE00059 pregnane_steroid_receptor	126.1	4.3e-37	2
CE00544	CE00544 estrogen_receptor	94.9	1.9e-27	2
CE00547	CE00547 steroid_receptor_Ad4BP	89.7	1.2e-26	1
CE00543	CE00543 steroid_receptor_N10	85.2	2.3e-36	2
CE00036	CE00036 androstane_receptor	69.3	5e-25	2
CE00208	CE00208 Progesterone_receptors	54.8	3.8e-15	2
CE00339	E00339 vitamin_D_receptor	43.7	3.4e-12	2
CE00546	CE00546 glucocorticoid_receptor	40.1	1.3e-10	2
CE00545	CE00545 progesteron_receptor	36.9	3.4e-11	2
CE00386	E00386 mineralocorticoid_receptor	10.2	0.31	1
PF00907	T-box	3.8	4	1
CE00341	E00341 seven-up_receptor	3.4	3	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00341	1/1	82	97 ..	175	190 ..	3.4	3
CE00545	1/2	81	148 ..	580	647 ..	25.6	8.9e-08
CE00546	1/2	81	148 ..	438	514 ..	26.8	5.9e-07
CE00208	1/2	81	148 ..	568	635 ..	47.1	5.6e-13
CE00386	1/1	88	149 ..	1	66 [.	10.2	0.31
CE00055	1/3	83	149 ..	109	174 ..	91.7	3.4e-31
CE00036	1/2	83	150 ..	21	88 ..	53.5	2.7e-19
CE00339	1/2	83	151 ..	32	100 ..	43.2	4.8e-12
CE00543	1/2	83	156 ..	327	400 ..	75.4	2.7e-32
CE00059	1/2	83	156 ..	39	113 ..	73.5	2e-21
PF00105	1/2	81	156 ..	1	77 [.]	190.7	4.3e-56
CE00544	1/2	83	157 ..	189	263 ..	75.4	9.8e-22
CE00547	1/1	83	157 ..	13	87 ..	89.7	1.2e-26
CE00342	1/2	83	200 ..	1	122 [.	158.3	2e-45
PF00907	1/1	212	228 ..	1	18 [.	3.8	4
CE00545	2/2	235	255 ..	745	765 ..	3.3	0.6
CE00546	2/2	231	256 ..	599	624 ..	2.6	3.1
CE00055	2/3	236	257 ..	297	318 ..	14.7	9e-05
CE00544	2/2	238	264 ..	372	398 ..	9.5	0.027
CE00036	2/2	237	264 ..	185	212 ..	12.5	0.00017
CE00208	2/2	226	266 ..	724	764 ..	3.8	1
CE00543	2/2	242	291 ..	494	543 ..	0.5	0.43
PF00105	2/2	326	334 ..	1	9 [.	5.4	2.7
CE00055	3/3	314	380 ..	376	442 ..	15.4	5.3e-05
PF00104	1/1	225	383 ..	1	167 [.]	241.4	1.3e-68
CE00059	2/2	235	414 ..	253	432 .]	47.4	1.2e-13
CE00342	2/2	223	416 ..	119	313 .]	176.7	8.7e-51
CE00339	2/2	323	417 ..	348	443 .]	-0.7	31
CE00550	1/1	55	457 .]	97	509 .]	1204.3	0

```

1   GTCCTTGGGT AGCATGTACA TTTCCATCCC TTCCTTTTAT ATATGGGGGT
51  AATAGGATAC CCCCTCCTCC AGGGGTATCC CCTCTTTCTA GGGACCTACC
101 CAAGCTAGGC CTTTCTTCCA GTGAAACGTG CATCCCAGAG GCTTCTAGGA
151 TGAAGTAGTC CACTGGAAGG CACCAGCTCT TCCTTTTATC TCTCCAGAGC
201 TGGACAGTGC ACCAGGGGCC GGTACTGGTT CCCCAGCTAG GAGACACCTT
251 GGGCGGGGCT TTGCTCGCCG GAAGCACGCA GAGCGTGGGG AGGAGGGCCC
301 CCTCTGCCTG TGTTGTGCC AACAGCACCC GCGCTGCCGC GTCGGGTTC
351 GGGCGCCGGA GTCACACATG ATGTCACAGA CAATGACACA AGCCGGTGTC
401 TCATTCCGAC ACAGCGTCCG AGCTGCACAA TGTCACACCC GGGTGCCAAA
451 CACTTGGCCC CGCGCGACCC GGCCCTACGC CTCCTGCCGC CGCTCTCCGC
501 GCTTCGCGGG CAGGTGGCCC GGTTCGGCCG GGCAGGGGGC TGGCGGGCGA
551 GCCCCGCGGG CGGGCTGGCG AGCGGGTGAT GTCACGGGCA GCGGTGGGTG
601 GGTCCTCGG AGGTGAGGCG CCGCCAGGCG AGTTCAGCGA GAGTTCAGCC
651 GCATTGCATT AGGCAAATGA GGCCCGGCTT GGGTGGGGGT GTGTGTTAAG
701 GGGAGGACAC CGGACCAACC CCCCTCTTCC CCGCCCCACC ACCTCCTCCA
751 CCACGGCTTC GCTCGGCCAG GGACTGACCA AACCTTGGGG GAGCCTGGGA
801 GCCGGAACGT GTACAAGGGG AGGACGCCCC CCCCTCTTCC GTCCTTGTC
851 CCTCGCAGCC CCCTCCTCTC CCTGTAATCG GCGTCCCTCT GACTCTGTG
901 TACTCCTCAT CTGGAGCCTT TCCCCCTTCC TGCTTCTCTC CTCTCCTCCC
951 CCTTCCAGG CTGCCCCCAC TTGCCTGTCC ACATGCCGCC TCTCCCTCTC
1001 GGTTCCTGCT GTTCTCTCCG CTGCAGCCGG ACGCGCCGGG AATGGGTAA
1051 GCCAGGGGCG GTGCCTGGAC GGGGCGGGGC GGTGGAAGG GGGTGGTGCC
1101 CGGAGGGGAG GGGGCGCGCA GAGCTGGGGT GGGGGGGCCG TGGCGCGTAC
1151 CACGAGAGAC CGAGCGATGC GCCAGCTGCC CTTGGCCTGG CGGGGGCGGA
1201 ACCGCGCGGG ATCCCCACCC CCACCCGGA TCCTCGCCAC GGAGAATCCC
1251 TGGAGAAGCC CCGGATCCCC GGCTGGGAGG AGGAAGTGCT CGTTGACCCC
1301 CAGCCCCGCG CTGATCCCGC CCGGGCCTG CGGACTTGGG GAGCCGCTGT
1351 ACTCTGCCTC GGACGCCACG AGACTCTAGA CGGGAGTCCC CTCGAGGTGA
1401 AGCCGCTGAG TTCCCGGGCC CCGCCAGGCT TCCCTGGGAG AGCCGACGGA
1451 CCCCCCTTCC CAGCACACAC AACTTCCCTG CTTTTACCG GGAAGGCGG
1501 AGCGCGCGCG GGACTTAGAC GCGGGGACTT CAGGGCAGGG GCGGCCCCCT
1551 GCGCGGGTCA CCAGTCGGGG CGAGGGGACG TCTCCTCTCC CCCAGCTGCT
1601 CTGCTCGGAT GCGCGCGCCG GCTGAGTGAC GGGGGCGGCG CGCAGGACTT
1651 CCCAGCTCGG ACCTCTTGCC TTCGAGGGGA AAGATGTACG AGAGTGTAGA
1701 AGTGGGGGGT CCCACCCCTA ATCCCTTCCT AGTGGTGGAT TTTTATAACC
1751 AGAACCGGGC CTGTTTGCTC CCAGAGAAGG GGCTCCCCGC CCCGGGTCCG
1801 TACTCCACCC CGCTCCGGAC TCCGCTTTGG AATGGCTCAA ACCACTGTAC
1851 GTACCGGCCT CTCAGTTAGC TGTGTAGGG GGTGGGAGTG GCGGTAGGG
1901 CTTCCTACTA TACTCGGGGG TGAGAGTCCC GGGGTGTAGT GGAGGTCTG
1951 TCTCTACCTT TACTTAACC CGTGTGCCC TTGCTGGACA ATTGAACCCT
2001 CCGCGCCGCA CCCTCCCCCC AGTAACCCTA AGTGCAATTT GTGTTAGATT
2051 AGGGCTGAGG AACTTTGAGA GTTCCTTCTT TTCAAGCAAC ATTCCTTTCA
2101 TCTCTTTGTT TCACTTCTTC CCAGGAGAAA TGAAGCCCAA GCCCCCTTTG
2151 GCCCCAGTT TGTATATTCT TTCTTGCCCT TGGGAAATCC CAAAAAGGTT
2201 TCACCCAGCA GGCTTGGGAA GGGGTGGGGG GGTAAAAGGG TTCCCTGGTC
2251 TTGTGGTGGG TTTTGGTCT TGCTTACCCG GGGGGGNNNN NNNNNNNNNN
2301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNGAAG GGCTCTGTGC
2351 AACTCAGGA GCTCGGAGCA CCAGGGTGTA CACCTGGGCA TTTTCTGCG
2401 CAGCTGTGAG GCACTGTACA CTGGGTGGGC GGGAGCAGGC GCAAGGGGGT
2451 TATTGTTAGA TGGCTCAGT TTCTTCCCCT CCTGGGCTTT GGGCTCTTTG
2501 CTGGAGGGGA AGCTCTTCCG TGGAGGATCT CCCACCTTCC TGGACCTGCT
2551 GCCTCCCTCC TGCTGCCAG GGAGGAGGGG TGGAGTGGGT CTCGGGGGGG
2601 CCCTGGCAGA TTGGAAAAGG TTGAAGGGCA AAGGACTTAC CCCACCCCTC
2651 TTGCTGGGAG AAGAGAGACC TGAGATGGAC AGACAGCCCA CCTCTGCCCT
2701 CCCAGAGCCA CTCTATCCC AGCTTTTCCT ATTGTCCTGC CCCCACCAT
2751 TTCTCTAGG GCCGAATCTG CTGTGTGGCT GTAGACACAA GAGGGAAGGT
2801 ATCACCCCTG ACTTTGGAAG AAGAGAGAGT GAGAGGATGA CTCTAGGACC
2851 CTTTTTCTCA TTCTCCAGT GCTGGAGCAA GACCCCCCTC CCTAGGGGG
2901 ATAGTTGGAG CAGGGCTGCC CAGAGTCACC CCTTCCACTG CCTTGGCCAC
2951 CTCTCCAGA GGGCTGGAGA GAAGCTGGGA TCTGAGACCT TGGTCTCCAG
3001 CCCCTGTCTC TTCTTAGCCC ATGGGGACAG CTCAGCTCTT CCTGGCCCAG
3051 AACTGGAGAG GGAGGAGGAT CACAGAGAGT AGGACAGGCA GTGTATTGGT
3101 GAGCCCTTCC CTAACCAC TGGACATGGG GAAGTGGAGA CCTGTCCCCA

```

FIGURE 3, page 1 of 8

```

3151 CATCCATTCT GGGGTGGGGT AGTAGACCTA GAGGCCTGGG TTTCCAGTTC
3201 CCGTAGTCTG AGCGTGGGTG TGCATATATA AGTGAGTGAG GTGTCA GTGG
3251 ACTCGGGTCC TGAGGCTGTG AGGTTGGGAG TGATGGGGGT CTGGGGGCTT
3301 GCCTTGAGGC ACAGGAAGGA CCCGGAGTCT GAGGGTGGCA ACTAGACTCA
3351 GTCTAGAATA TGTGGGGCCA ATGCCACCAC CTTGGAAGGG TCCCCTTGGG
3401 TGTGTTGGAA GTCCGCTGGT GACTGGAGCT GCCTCCAGCC CCCTCTTGGG
3451 GAATTCTCCA CTCTCCCCTT TACTGCCACT GAAGGTGGGA AGAGCAGGTT
3501 GGCTCTGGGA GGAGGTGGCC TGGGTTCTGC AGGGCCCTAG GGACATTGCC
3551 TCCCCTCCCA GAGCCCTCAT TTCGGTGCAT TAGAGGACAA GGGGGGGTGC
3601 ACAGGATGTG GCTCCCCATC TGTCTCCCAC CAATCTCCGC CACTCACACC
3651 TCCGCCCGCT CCCAGACGTC CAAGAATGTG AAGCACGTGG ATGCCCGTAG
3701 TTGGGGGAGG GGGAGACGCT TATCAGGCGG CCGCTGGGCT AGGGGCCTTC
3751 TTCCGCTGCC GCGGTACACC CAGAGCTACC CCCGCTCTC CCCGGGAGGA
3801 GGAAGGACGG TACAGAGGGC CCTACGCCCC CTCCCCAACC ATCCCCAGGG
3851 GCTGCGAGGG GAGCTGCGGA GGAGCGGGCG CCAGCTGGAT TGGGAGGGGA
3901 GCCGCTGGCC GGGGGCCCCG CTGATTTCTT GCTGATCTCC TCCAGGAAAC
3951 GTGGCCCTTG TCGAGCCTG CGAACGGCTC GGGGGCGTGG GGAATCCGGA
4001 CTGGAGCGCT CTGCGCCGCC CGCCTGCCA GGATGGGGAG CGAGGGAGGG
4051 GCACCTGGC AGCGTCGGCG GGAGGGGACG CCTGGCTTCC TGGGTCA GTT
4101 CCAGTCCTCT GTTGGGCGCT GGAAC TTTGA GCTGAGAAGG TGTGGTCTTT
4151 CTCTAGCCCG AGTCCTTCTG CAGGAAGAGG AGAGATTGGT GGGCTGGGCC
4201 TCTGGGAGG GAGGTTAGCA GGGATGGGCC AGGCCCGGGC AGTCCCCTCCC
4251 CCGTTGGTGT CCTCCCCAC TCCACCTGTG TGTGCAGGGA GTTATGGCCG
4301 TGTCTAACT CTGTCAGAGG CTGTGAGGAT TCCGGAGTTC CCCACACCTC
4351 CGGCCTTGGT CCTGTACCT CACCTCCTTG GACTGCTGGC TGGAGGCCTG
4401 GGGAGGTGGG GCATCGAGCT CTGGGTTCAG AGGGCAGAGC AGGGAAACCT
4451 CAGAGCTGGG TTACCTGGGT GACAGGTGGG GATGTGCTGG AGGTAGGGGG
4501 CAGGCTATGT TACAGCCTCC AAGGCAGTCA AGCTGCCGTT GGGTGGGCTA
4551 AAAGGAGGCC TTGCCCAGCC TAACTGTAG TCCTTGCCCTC TGGTCATCTC
4601 TCCCATTCTG CAAAAAATA ATTTTAAAA GCACATTCTC TCAGTTCGGT
4651 AAACACCCTC TGTGGAGCT TGCTTTAGCT CCATGTTTTT ATGGCTTTTT
4701 GCCCTCTAGT CTGTCCAGG CCTTAGAGCT GTTTACCTCT CATCCTGGTA
4751 TCCCCATGA CTCCCCATAC CCTAGCTCCC CTCGTGACAT CCCGCTCTGT
4801 ACCCCCAAAG CTCCCTCAGT CCTTCTCCC TCTCCAGTCT GGTTCATTTT
4851 AGAAGTGGGG CCTTGGGAGA GGCGGGGCCC AGGGCAAACG GTGGATTAGG
4901 AGGGGTGGGG AGGTCAGTGC CTCTTCTCTC TGCTTGTCGG AATGCTGACC
4951 AAGATTCTAG GCCATGGTCC CCCCACCCCT CCACATACCC CCTTGCCCTT
5001 GATCTCCCTT CCCCCACCA GTCTGGATTG TCTATTGTTA CTGCTTTTAC
5051 GTCTTGAAA AAGTTAGCAC AACAAAGGGC TGCTTTGTGG CTCACCCCTC
5101 CTGCCCTCTG GCCTCACCCA GGCCCCCAA CCCC GCCCC CCAGCAGCTG
5151 TTCTCAGGCC TCTAGCCTG TCTGATTTGC TTGTCTGGCC TGGGGAGAAT
5201 GAGGTGGGAG AAAACCAGGC CAGGCGAGTT GGTGTTGGAG TGAAGAGCAG
5251 ACGGCGGTGG GGAGGTCAGG AGAGAATCTG CTGGGCTGGG GATGGTGTGG
5301 GCATCAACTG TCCATTGCT GCAGGCTGGT CTTGGGGCAG GGAAGGGGAT
5351 GGGGGGCCAT AGCAGTGCTG GTCAGCCAGG CTGGCCTGGG AAGTGGTGCC
5401 CAGGCACTAC TAAGAGCCAG GAAAGCCCTG CCAAGGTTGT TGGCCTAGTT
5451 CCCTGTCATC AGCCGCTAG CAGCCCCAC TGTGTCTGCA GGTAAGGGGG
5501 GAGGTTGGTA GCACATAGTC AGCCCTGGT GTTCCCATGC TTCCTTCTCTC
5551 TGTGCCCAA TTTTAGGGCC ATGTGATTTG GGGCTATGTG ACTCATGTCT
5601 GTAAGTGCTG TGGGCCAGGA GCTGTGGGCA CCTTTAAATG CCAGCCAGTC
5651 TCATGTGCCG GAGTTGGGG TAGGGCTAGG TAGGATTGTG GAATATGGGA
5701 GGAGGCAGGG ATCTGTCTAC CTAGGGAGGC ATCCTCATCC ATCCTTGGCC
5751 CTGGACAAGA GAACTGAAC GTTGGTAGGG GCCTCAGGAC GATGCTGCGT
5801 GGCCCTTGG GAATCTGGGA TTGTCTGGT CATAGTTCTT ATCTTGACCC
5851 CAACACCCTT AGCTGCCAG GCTTTGGACA TGATAGCCC CTACCCAACC
5901 CAGCCCTGTT CTGCCACAG TGATGGGCAT GGAGCCAGAC ACTGGGGAGG
5951 ATTTGGCCAG TGAGGGCTGC CCCTGCTGTC TGGGTCACCC CTCCTGGCTG
6001 CCTCTTGA GCTGAATAAC AGAAGGGGAG GGTTAGTAA CCCGGACATA
6051 GTATTGAGGC CAGACAGACA GAGCATTGAT GGAACAGAC CCCCTTTGTC
6101 ATGCCATCTC TCCCAGATG GGGGTACCC AGAATAATGG GCTTTTGGGG
6151 CCCTGGGGAC TCTTCTCCCT GTATTAGGG TATCTCCCC TATCTCAGGG
6201 AGACACCTCC TACTGTGCCC AGCATTTGTG ACTCTTCTTT GCACCCCTG
6251 CCTTGGGTCC CTGGCCCTGG GATTGTTTGG GTGGAGGAGG GGCAGTGGCT

```

FIGURE 3, page 2 of 8

```

6301 GCTGGCAGAA TGGGGTGGAG GGGGGAGCGG AAGCAGAGGG GGCGGGGGAG
6351 TGGCCGGCTT TGAATATCCT GTTGACCCCA GTTTCCTCTG CCCCAGCTT
6401 ATGTCCTCTT CCCTCCCTCC TCTTCAAGCG TTAATCCTT CCTAACTCGG
6451 GGGGAGAACG GGGCCAGGCC GCCCAGGGG AGGAGCTTTA GAATCAGGGT
6501 GACCCCCACC CCTACTCCCC AAGCACAGTC ACGGCACACA TACAAATGTG
6551 ATGGTTTATC ATTGTATCTT TGTGGTTTTG AAGGTGGGGG TCCTAGGAGT
6601 CCAGAGGAGT GATGGGGTGC TGGAGGCTTC ATTGGCAGCC TCCTGCCCTG
6651 AGTCTGGCTG GGGAGTCCCA GTTTTCTTAA GACTTGAATC CTGCCAGCAG
6701 TGGTGAGGCT GGGAGAGGCT CTAGGAGGG ACGGTGAGGC AGGGTGGAGC
6751 TTGGTACTAA GGATGGCGAC CTAGGTCTCT AACTGCCCCT CCCCTCTTCT
6801 TCTCTAGACC ATTGAGACCC AGAGCAGCAG TTCTGAAGAG ATAGTGCCCA
6851 GCCCTCCCTC GCCACCCCTT CTACCCCGCA TCTACAAGCC TTGCTTTGTC
6901 TGTCAGGACA AGTCTCAGG CTACCACTAT GGGGTGAGCG CCTGTGAGGG
6951 CTGCAAGGTG AGTTGAAGGG GTCATTGGGA AGGACAGCTT GATGAGGTCA
7001 ATGGGGATGT CCCCCTTCT GTGTCTGGG AGTGTGCAGT TGGGGGGTGT
7051 CCCTGAATTG CTGCTCTTCT TTCTCTGTGG AAGTTGGCAG CAAGCAGGGG
7101 ACACCTACCA CAGTTTCCCC ACAGGTCTCT CCCCATAAAT GTGCAGGGCT
7151 CCCTCAAACC AGAGGTCCCC TCCTGCCTCA GCTCCTTTCC CTGTCTCTAT
7201 CCTCCAGCTG GCAGGGCGTA CGCTGCTCT GCCACCGCTG CCCAGGTTGC
7251 CATGGTGAGC TGGCTGCCGA CTGGCTCTTG GCTGGGGACC CAGGAGGCCT
7301 CCCCAGGCGG CCCTGCCTGA ACCTCACCAT GGAGCCTGG CAGGAGGCAG
7351 TTAGGAGCAG GCACCTGCC TTAGCTTCCC CTTCAGGTGC CCGGGCTGTG
7401 GGCTCCCCAT TGCTGGCTG GATTTCCCCA TCCTCACGTT AGGTGCCAGG
7451 TGCCAGGTAT ACCTGGTCTT TAGCAGCCCT GCGCCCGGCT TCTCCTCCTT
7501 TCCCTGGGGC CTGAGCCTCT GTGTGCGTTT CTTCCTCCAG AGATTGGGGC
7551 TCAGAATCTT CACAGCTTTG GGCTTGCAG CTCTGGGCTG CTCTTCAGCC
7601 TGGAGTAGCT ATCCCAGAT GTGGGACGGA GGTCAAGGGC AAAGCACAAG
7651 GACTCAGGCT GTGTGTCTGC CTGTCTGTG TGGTTGTTCC TGGTCTGTTC
7701 TTCTCTGTG CGCTGTCCC TCTGGTCAGC CTGTATGTGG AGCCCCCTGGC
7751 CAGCCTGGGT CTGTGTCTGT GATGGGTCGG TGCACACCTG TCTTGGTGAA
7801 CTCACATCTT TCTGCCTGT TCCTGAGTGC ATGTGTGTGT TCGCCTCCAT
7851 TTCTCTGGCC AGCCCGTGTA TCTGCCTCCT GGCTCTTCG GGCTTGTCTT
7901 CTTTTCTGT GTTCTGAGT CAGGGGTGTG GGTTCCAGAT CCCTGGCTGT
7951 TGCCAGTTA GCCCATGTC TTCTATTTT TGAATCACC GCAGCCCTGA
8001 GGTCTTTTCC CTGGAAGGGA GGAGTCAGGT GTGTGCTGTG GGTGGGGGA
8051 AGACTCCTGC CCATCCTGCA GTGTTGAGGC AGGTACTGGG ATTCTCCTGA
8101 GGAGGATCCT TTTAGGTGAA TCATTCTCCC CAGCTTTTCT GGCTGTCTCA
8151 GGTAGCGCAT GGGCAAACGC TTGGGGGACG CAGCTGGCCT GGCCCTCCTC
8201 CCCTAGACTG AGACCGTAGC CAGGCACTGC TCCCACTGTG GGTGTGGACA
8251 ACCTGACTCC CTCCCCTCCA TACCCAGGGC TTCTTCCGCC GCAGCATCCA
8301 GAAGAACATG GTGTACACGT GTCACCGGGA CAAGAACTGC ATCATCAACA
8351 AGGTGACCCG GAACCGCTGC CAGTACTGCC GACTGCAGAA GTGCTTTGAA
8401 TTGGGCATGT CCAAGGAGTG TGAGTGCCAT AGGGCAGGGG CCGAGTCCCG
8451 CCTCAGTTGG GGTCTCAGT GCTCCTAAAG ACCAAGGGAG CAGGGCTCTG
8501 TGGATGTTTG TGACATGCA TGAACACGCA TGCCGTGGTG TGCGGGCTCA
8551 CGGTTGAGGA TGGTTTGTGT GTAGCTGCAA GGACCTGTTT GCGAGTCTGG
8601 CTGGCTGTGT GTCCACGGGC AGGTCTGTGC TCCGGGACCG TGTATGTGTA
8651 ACCATTCTCTG TTTCTGCACG TCTGGCTGTG TGTGCTTGCG TATGTGTGTG
8701 TGTGTGCATG CTCCAGGATG GCTTTCCTCC AGGCCGTGCT TGGTTTTGGG
8751 TTGGGGCTCA GAGGCATAG CAGTCCCTTC TGATTGTGAG TCTTAGGGGA
8801 GGGGCTTGAA TTCTGAGGGG TGCTTGGCTG GACTTATGTG TGTATGGGGG
8851 GGTGGAAGGG CTGGCACAAG GATCCAAAAG CCATTGTCTA GTTAAGCCTG
8901 GGATTGAGAG TTGGAAGAAA GAATTGGGAC TTCTCAGATC CCAGAGGAAA
8951 CGGGGTTTCC ACTTTGGGCT CAGCTGAGGC CTGATGGAGG GAGGGAGGGA
9001 AAGGCTGGAC AGGGAGACCC TCTTGTGTTG AATCATGGGT GTTGCCATGG
9051 TGACCGGTGA TTGATGATGT CAGAGATAAA TGACGCTGAC AGACGCCTCC
9101 TTGTCTGCGT GGCGTTGCC ATGGAGCCTG AGCCTTGGGG GATGGGATGG
9151 TGGAGGGGGC TGACGAGCCC CCTAGCCCTT TGTGGGGAGG GCAGTGGGGA
9201 GGGGGCACGG GTGAGATGGT TCTGACTGTT GCACGAAGAG CCCCAGACAG
9251 GAATGGAGGG GACTGGAGTG TCCTGCCACA GGAGGCTGGG GGTGCCTTGT
9301 CCTGAGCCCA GGAAGTGGTG GCTCCTGCTG CAAGAGTGGG TGACAACCTCA
9351 AGACCCACAA GCCTGGAACC CTTCGCTTAA GGGCTGTCAC CTCCTCCTCT
9401 CTGTTTGTGC CACCTTCTGC TCTTTTCATG GCAGAAGGAC CAGGGAGGGG

```

FIGURE 3, page 3 of 8

9451	ACCCCTTCTC	CCTCCCACCG	CCAACTCCCC	TTCTCCCTCC	CACCGCCAAC
9501	TCCCCCTCTC	CCGGCTGCTC	TGTGCCCCGG	AGCTGAGCAG	CTGCCATTTT
9551	AATAGAATTA	AAGCTTCCGA	ATGATAAACG	TCTTGTACAC	GCTGCAATTT
9601	TCTCTTCCCA	AATTATCCCC	CCACTCTCCC	TCTCCCTCTC	CCTTCTCTCC
9651	CCTGGCACTT	ATTGAATTTG	CAGAATCGAC	ATGAGTGATC	TCCAAATTAT
9701	GCCAGCTACC	CCCACCTCGC	TACCCCTCC	CTGAGCCCCT	CCCCACCCT
9751	CCCTTCCTCC	CGGTCAGCA	GCCACCACCA	CCAGCCCTGT	GAGTGATTGT
9801	GTGTCTGGAT	AATCGGCTGG	TAACGACCCC	ATCGCTTCTT	TAAAGCCGAG
9851	TGGTGTGTGC	GGCTCAGCGC	CCCTGGTGAT	TTGTCAGCTC	CCCAGCTAAT
9901	GGGCCAAGAG	ATTCTCCCCG	CCAGGTCCCC	CACTCTCAGG	CTGGGGAGCC
9951	CTACTCCCCA	CTTGCCCCAG	GAGCTGCTCA	GAGCCAGTCC	CAAGGGACCC
10001	CCATGGGAGAC	TGCAGCTGGG	AGGGCTGGGT	GAGTGGAGGC	GGGAGAAGGA
10051	CCTTCCTGGG	GAAAGAGGAG	GCAGAGCACC	TAGGAGGGCA	CCGTGCGCTG
10101	GAGTGTGAGC	TGGAGTAGAC	GCGTGGGGGA	TAGCATGCGG	CTGGCTATGG
10151	GGTGGGGTGG	GGGGTGTGTG	CAGGGCCACA	GCTGTGCTCA	TGGGGCTTCT
10201	GGGGCAGAAC	TTGATGTGTG	GGTTGGGTGG	GCATGGAGGG	CTGGAGTGCG
10251	TGGCAATGCC	TTGCCTGCCC	GTGAACGCGT	GCTGTGTGCG	CGTGCTTACA
10301	AGCCTGGGTG	ACCTCCTCAG	CAGCTGGCAG	CTCTCTGTCA	GGCTGGGGGT
10351	GGACGAGGCC	CTGAGCAGCC	TGCAGCTGCC	CTCTTAACCC	CCTCTGCCCT
10401	CCACAGCTGT	GAGAAACGAC	CGAAACAAGA	AGAAGAAGGA	GGTGCCCAAG
10451	CCCGAGTGCT	CTGAGAGCTA	CACGCTGACG	CCGGAGGTGG	GGGAGCTCAT
10501	TGAGAAGGTG	CGCAAAGCGC	ACCAGGAAAC	CTTCCCTGCC	CTCTGCCAGC
10551	TGGGCAATA	CACTACGGTA	TGGCTTTCCC	CCGGCCTGCA	GGGTGGGATT
10601	TGCCCAGGGC	CACAGGGCCA	GGATGGGCCC	CTCTCAGGCA	CCCCTTCTTG
10651	TGCCAGGCAA	GATCTCTGCG	TCCTTCCCTT	CCCCTCTCTT	CTCCCTCCTC
10701	CTGCTGCCTC	TTCCCAAGGA	GCTCCCAGGA	AGTGAAGGCT	GGGTAGAGGG
10751	CAGGCTGTG	GGGGCTGGAG	CCAGGCTGAG	AAGGGGTGCC	ATGGAGAAGA
10801	AGGCCCTCAC	TCTCCCTCCT	CCCCCAGAAC	AACAGCTCAG	AACAACGTGT
10851	CTCTCTGGAC	ATTGACCTCT	GGGACAAGTT	CAGTGAATC	TCCACCAAGT
10901	GCATCATTA	GACTGTGGAG	TTGCGCAAGC	AGCTGCCCCG	CTTACCACC
10951	CTCACCATCG	CCGACCAGAT	CACCTCCTC	AAGGCTGCC	GCCTGGACAT
11001	CCTGGTGAGG	GTCTGCACCC	TGGCCCCCAG	GCACTGCCCC	TGTGTCCTGG
11051	GTAGATGTCC	TTCCAGCCAG	ACAGCCACCC	TCCTAAATGT	CTGTCTGCAA
11101	TCAACCTGTC	CAAAAGCCCA	CCGCCCCAAT	GTCTGCCCTT	CCTCTCCCCA
11151	TATGTCCACC	TGTCCACTCG	TCTCCCTGTC	CACTCAGCCA	CCTAGCAGCC
11201	AGATGTGAC	GAGCTCACCT	GTTCACCCAT	ACACATATCC	AGCCACCCAG
11251	CCATCCATCC	ATTTAGCCAG	TAATAAAGAT	TCACGTAGGA	GCCAGGTGCA
11301	GTGGCTCATA	CCTGTAATCC	CAGCACTTTG	GGAGGCCGAG	CGAGGCAGGA
11351	GGATCACTTG	AGGCTGGAAG	TTCAAGACCA	CCCTGGGCAA	CATAGTGAGA
11401	CCTTATTTCT	GCAAAAACT	AAAAAGATT	ACCTAGGATC	CTCTGGCCAG
11451	TGTTGAGCT	GGGTGTCAGG	AACCCAGCGG	TGAATGCACC	ACCATCCCC
11501	CTCTTGAAAA	CCTTCCATGT	GAGGCAAGAG	ATAAGTCAAC	AGAGGTTGCA
11551	AAACTGTGAT	CAATGCTTCC	TGGAGATTGG	GGGAGGGCTT	GTGACTGCTT
11601	GGGCCTGAAG	GATGATGTCT	CAGAGGAGGT	GACATCTAGG	GGTTTGTA
11651	GGGGGAGGTG	AGAGGGTAGC	CCTAACTCAG	GAGCAGGAAG	TGAAAGACTT
11701	GCTGCTGTGA	GGCCATGCTG	AGCTCAGGGG	ACTGCCGGGC	ACTCGGTGAG
11751	GTGAGCCCGA	GGGTAGACTG	GGCTGGAGGC	TGGATGCAGG	GGGTGGGGGC
11801	AGGAAGAGGT	GGTGGGAACT	GCCAAAGCCT	AGGCTGGAGG	GAGCACTCTC
11851	CTTCTGCTG	TCCTTGACAA	GGGCTCGGTC	CACCTGTTCC	CTCTTGGTCA
11901	CCTCCAGGGT	GGGGAACCTG	GGATTGACG	AGACTGTCT	TTCTTTTTAT
11951	GTTTTTCTTT	TTTGAGATGG	AGTTTCACTC	TTGTCACCCA	GGCTGGAGTG
12001	CAGTAGTATG	ATCTTGGCTC	ACTGCAGCCT	GCAACTGCTG	CCTCCCGGGT
12051	TCAAGCGATT	CTCCTGCCTC	AGCCTCCTGA	GTAGCTGGGA	TTACAGGCAC
12101	CCGCCACCAC	ACCCGGCTAA	TTTTTGATT	TTTGTAGAGA	CGGGGTTTCA
12151	CCATGTTGGC	CAGGCCGGTC	TCGAACTCCT	GACCTCAGGT	GATCCTCCCG
12201	CGTGAGCCGG	CAGACTGTCA	TTTCTCCATG	GGCACCTCTG	AATGTTGAGG
12251	CGGGTGATGG	GTGGGAGGTT	TAGATTGTGC	TGCCTGCAGG	GGCTCCCATC
12301	CCCATGCCGT	GGATGCAGGA	GGTGCCGTCT	GGGTTCTGTC	AACCATATTC
12351	AAGCCAATAC	ACATTTACTG	AGCGCTTGTT	GTGTACCTCA	TCCTGGGAGC
12401	TGTAGGCAGC	AGCCCAGTGT	TCCTTAGCTC	CTAGAAATTC	TAGGTCCCCT
12451	CTACATTCTT	TGCATGTAGG	CAGGATGACC	TGGACCTGCA	CTATCCAGTA
12501	CAGTAGCTGC	TCACCACATG	TGACTCTTTA	AATTTAAATT	AATTTAAATT
12551	AAACTCAATT	CAGTTCCTCA	GTTGCATTAG	CCACATTTCA	AGTACTCAGT

FIGURE 3, page 4 of 8

12601	AGACGCATGT	GGCTGGTGGC	TGAGGTATGG	ATGGTGCAGA	CGTAGAACCT
12651	TTCCATCATT	GTAGAAAATT	CTATCAGACA	GCATTGCTCC	GGCCACCTGC
12701	CAGGTGGTCC	TCCGGGAGTG	CTGGTGCGGA	GTGCTGGTGC	CGAGTGCTCA
12751	GAGTGGGTTT	GGGTTTCAGT	CCTGAACCCA	AGCATCCTCT	GCACCCAGAT
12801	CGTGC GGATC	TGCACGCGGT	ACACGCCCGA	GCAGGACACC	ATGACCTTCT
12851	CGGACGGGCT	GACCCTGAAC	CGGACCCAGA	TGCACAACGC	TGGCTTCGGC
12901	CCCCTACCG	ACCTGGTCTT	TGCCTTCGCC	AACCAGCTGC	TGCCCCTGGA
12951	GATGGATGAT	GCGGAGACGG	GGCTGCTCAG	CGCCATCTGC	CTCATCTGCG
13001	GAGGTGGGCA	GGGGGCCTGG	GTCTGGGGGC	TGGGCTGGGA	CGGGGGTGCA
13051	GCCCTGGAGT	CTCTTCCAGG	GAGCTCTTTC	AGGCCACCTC	TGTTAGGTAT
13101	CTCTAGAGGT	CAGGCTCTGG	TCTGCAACTA	CACAGCAAGG	GGGCCATGTG
13151	GGCGCTGGAC	TCTGTTCCTC	GATTTCTGGG	CAACACCCCT	TCTAGGGAGG
13201	TTAAGAGTGA	GGGTTTGAGG	GTCGGACCAA	CCAGGGTCAC	CTCCTGGCCG
13251	ATGCATGACC	CTGAGCAGGT	TGCTGAACTT	CTCTGGGCCT	CCGTTTCTGT
13301	ACAGTGGGGG	CGGTAACGGT	CTCTAGCTCA	TGAAGTTGAT	GGGAGGATTA
13351	CGGTGGTAAC	AGATACTGTG	CAGGTGCCCA	GAGCGAGCTC	CAGTGCTTGT
13401	TAGTTGCTAT	TTTATTGTTG	TGATTTCTGC	CATTTTCATCT	GGTTTCCAGA
13451	ATAACAGGGG	GGAGTGGGAG	CCTGCCTGGG	AACCTCTCC	CTGCTTGAGG
13501	ATGGCACTGC	CCATTTGGGG	TCCCATCCCA	CTAACTGGGC	TCAGGGAGGG
13551	TTTGGGGCAC	CCCCTACCC	TCAGCTCCCG	TTGCTCCCTT	TTAAGGGCCT
13601	CTGTACCCTG	CGGCAGCAGA	GACCCCATGC	CCTGCCCTGT	GTGGGGAGGC
13651	GCCTGCGAGC	TGCCCTCCTC	CATGGCCTGG	GCAGGCACGC	CCCCCGGTGG
13701	CCGAGGCTGG	GGGTGCAGCT	GTGTTCCAG	CTGCTCAGGG	GGTGGTTCTG
13751	CTTCCTCAGA	CCGCCAGGAC	CTGGAGCAGC	CGGACCGGGT	GGACATGCTG
13801	CAGGAGCCGC	TGCTGGAGGC	GCTAAAGGTC	TACGTGCGGA	AGCGGAGGCC
13851	CAGCCGCCCC	CACATGTTCC	CCAAGATGCT	AATGAAGATT	ACTGACCTGC
13901	GAAGCATCAG	CGCCAAGGGT	GAGGCTCACA	GACCTGGAGG	GGTACCGGCC
13951	CCCGACACCT	GGCCCAGGCC	CCCACATCCA	AGCCAGCACC	CCATGTCTTT
14001	GTGCCAGGAC	AATACGACAC	CTGTCCCCAT	CTGTGTCTAG	GCTGAGGTCC
14051	CCTAGTGACT	CCACTTTGCC	GAGGTGGCCC	GCCTGTGTCA	CCTTTGTGTG
14101	GTAGTTTCAGA	TCGTGGCTCT	GGAACCAGAC	ACGTGGGTGT	GTGTCCTTGT
14151	GTGGGTCACT	CAACAGCTCC	TAGCTACAGT	TTCCCTTCCG	AGGGCGGGGA
14201	TAACATTTCGT	GTTTACAGAG	GGGTGCGGAT	GATCCCTAGC	ACACAGCACA
14251	GGGAAGGAA	GGGCTTGCG	TCTAGCCCAG	GCCGGCAGTC	TGGCCCTGGA
14301	GCCGGAGTTC	GGGACCACTT	TGCCCCATTG	CCACCAGCCT	CTGGACCTGG
14351	GGGCTTAAGA	GAGCTGGCTC	GTGTCAAAGA	ACTGAATCCC	AAGAAAGATG
14401	CTAATATCAG	CAGTATTGAT	CTTCCCACCT	CGAGCCAGGC	TTGCTGGGGC
14451	TGGGGGTGGG	AGGGCTGGCC	CAGCGTGCTG	ACCTCTGCC	CCTCCTTTCC
14501	TGCAGGGGCT	GAGCGGGTGA	TCACGCTGAA	GATGGAGATC	CCGGGCTCCA
14551	TGCCGCCTCT	CATCCAGGAA	ATGTTGGAGA	ACTCAGAGGG	CCTGGACACT
14601	CTGAGCGGAC	AGCCGGGGGG	TGGGGGGGCG	GGACGGTGGT	GGCCTTCTGC
14651	AGTAAAAAGT	GCCCTGATGC	CACCATTGCC	GTAAAAACTA	ATGCCCAATT
14701	GTGATAAGGA	GCTACCGGGG	TACACACGGG	GACTGGTTCA	AATGGGGCAT
14751	CGCCGAAGCA	TGTGATGCTA	TGAACTTAAT	CGGACTATTA	TCTGTTGGA
14801	TCTCTAAACC	AGCATCGCAA	CCTGGACACT	CTTTTGCATG	GTCGTTATTA
14851	TCTCCGGTAG	ACTCCTTGCC	TCCCTTTTAC	ATAAAAAGGC	CTCCCCGAC
14901	AAAAAGGGTC	AGTTCGATCC	CCACTTTCGG	TTCGGGAGCC	TACCGTGTGC
14951	CAAAGGCCCT	TAATCTCGAA	AATATCCCAA	TTACCTGATG	TCGTGCGACG
15001	CCTAAAAATT	CCCCGTGTTG	CCACCACTGC	TTGAAACCCC	CAAGCTTGGG
15051	TGTTAATCCC	GAATTGGGGG	CCCCCGTNN	NNNNNNNNNN	NNNNNNNNNN
15101	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNG	GCCCCCCCGC	CAGGCAGCTG
15151	TAGCCCCAGC	CTCAGCCCCA	GCTCCAACAG	AAGCAGCCCG	GCCACCCACT
15201	CCCCGTGACC	GCCCACGCCA	CATGGACACA	GCCCTCGCCC	TCCGCCCCGG
15251	CTTTTCTCTG	CCTTTCTACC	GACCATGTGA	CCCCGCACCA	GCCCTGCCCC
15301	CACCTGCCCT	CCCGGGCAGT	ACTGGGGACC	TTCCCTGGGG	GACGGGGAGG
15351	GAGGAGGCAG	CGACTCCTTG	GACAGAGGCC	TGGGCCCTCA	GTGGACTGCC
15401	TGCTCCACAC	GCCTGGGCTG	ACGTCAGAGG	CCGAGGCCAG	GAAGTGAAGT
15451	AGGCCCTTGG	TCCTGGGTCT	CAGGATGGGT	CCTGGGGGCC	TCGTGTTTAT
15501	CAAGACACCC	CTCTGCCCAG	CTCACCACAT	CTTCATCACC	AGCAAACGCC
15551	AGGACTTGGC	TCCCCCATCC	TCAGAACTCA	CAAGCCATTG	CTCCCCAGCT
15601	GGGAACCTC	AACTTCCCCC	CTGCCTCGGT	TGGTGACAGA	GGGGGTGGGA
15651	CAGGGGCGGG	GGGTCCCCC	TGTACATACC	CTGCCATACC	AACCCAGGT
15701	ATTAATTCTC	GCTGGTTTTG	TTTTTATTTT	AATTTTTTTT	TTTTGATTTT

FIGURE 3, page 5 of 8

```

15751 TTTAATAAGA ATTTTCATTT TAAGCACATT TATACTGAAG GAATTTGTGC
15801 TGTGTATTGG GGGGAGCTGG ATCCAGAGCT GGAGGGGGTG GGTCCGGGGG
15851 AGGGAGTGGC TCGGAAGGGG CCCCCACTCT CCTTTCATGT CCCTGTGCCC
15901 CCCAGTTCTC CTCCTCAGCC TTTTCCTCCT CAGTTTTTCTC TTTAAAACTG
15951 TGAAGTACTA ACTTTCCAAG GCCTGCCTTC CCCTCCCTCC CACTGGAGAA
16001 GCCGCCAGCC CCTTTCTCCC TCTGCCTGAC CACTGGGTGT GGACGGTGTG
16051 GGGCAGCCCT GAAAGGACAG GCTCCTGGCC TTGGCACTTG CCTGCACCCA
16101 CCATGAGGCA TGGAGCAGGG CAGAGCAAGG GCCCCGGGAC AGAGTTTTTC
16151 CAGACCTGGC TCCTCGGCAG AGCTGCCTCC CGTCAGGGCC CACATCATCT
16201 AGGCTCCCCA GCCCCACTG TGAAGGGGCT GGCCAGGGGC CCGAGCTGCC
16251 CCCACCCCGG GCCTCAGCCA CCAGCACCCC CATAGGGCCC CCAGACACCA
16301 CACACATGCG CGTGCGCACA CACACAAACA CACACACACT GGACAGTAGA
16351 TGGGCCGACA CACACTTGGC CCGAGTTCCT CCATTTCCCT GGCCTGCCCC
16401 CCACCCCCAA CCTGTCCCAC CCCCGTGCCC CTTCTTACC CCGCAGGACG
16451 GGCTTACAGG GGGGTCTCCC CTCACCCCTG CACCCCCAGC TGGGGGAGCT
16501 GGCTCTGCCC CGACCTCCTT CACCAGGGGT TGGGGCCCCT TCCCCTGGAG
16551 CCCGTGGGTG CACCTGTTAC TGTGGGGCTT TCCACTGAGA TCTACTGGAT
16601 AAAGAATAAA GTTCTATTTA TTCTACACAT GCCTCCAGCC TTGCTGCCTC
16651 CACCCCCTCC TCTTGGCGTC TGGTCTGGGG GCTTGGGATG GGTTCGTCA
16701 TGTGCTCTGG GCCTGTGATG GCCAGGAATG AGCACTGGGG CCAAGGGGCT
16751 GGCCAGGGCA CCCTCCAAG CTGCCTTCTG AGGCTTACCT TGTGCTGGGG
16801 TCTTTGGAGA TGCTGAGAAG GAGAAAGTCC TGCCCCTTGG GAAGCCCTCA
16851 GTCTGGGGAT CCACACTGCC CATGTCAAGG AGCCCCAGTC TGGGAGTGGG
16901 AGAGAAGAGG AGGAAAGCTG CCCCCACCTT CAGGGAACCC CCAGTCTGAG
16951 GGAGGAAGCC GGAGCCACCC CTAGACATTT CTGGTCCTTG GGAAGCCTTC

```

FEATURES:

```

Start:      1684
Exon:       1684-1846
Intron:     1847-6808
Exon:       6809-6957
Intron:     6958-8277
Exon:       8278-8419
Intron:     8420-10406
Exon:       10407-10567
Intron:     10568-10827
Exon:       10828-11004
Intron:     11005-12798
Exon:       12799-13003
Intron:     13004-13759
Exon:       13760-13918
Intron:     13919-14505
Exon:       14506-14658
Intron:     14659-15143
Exon:       15144-15208
Stop:       15206

```

Map:

```

Bac Accession AC018629
Homo sapiens chromosome 17

```

SNP's:

Position	Allele 1	Allele 2	Context
4084	C	G	GCCTCTCCCCGGGAGGAGGAAGGACGGTACAGAGGGCCCTACGCCCCCTCCCCAACCATCCCCAGG GGCTGCGAGGGGAGCTGCGGAGGAGCGGGCGCCAGCTGGATTGGGAGGGGAGCCGCTGGCCGGGGG CCCCGGCTGATTTCCCTGCTGATCTCCTCCAGGAAACCGGCCCTTGTGCGAGCCTGCGAACGGCTCG GGGGCGTGGGGAATCCGGAGTGGAGCGCTCTGCGCCGCCCGCCCTGCCAGGATGGGGAGCGAGGGA GGGGCACCCCTGGCAGCGTGGCGGGAGGGGACGCCT [C, G] GCTTCTGGGTCACTTCCAGTCCCTGTTGGGCGCTGGAACCTTTAGAGCTGAGAAGGTGTGGTCCTT CTCTAGCCCGAGTCCCTTCTGCAGGAAGAGGAGAGATTGGTGGGCTGGGCCTCTGGGGAGGGAGGTT AGCAGGGATGGGCCAGGCCCGGGCAGTCCCTCCCCCGTTGGTGTCCCTCCCCACTCCACCTGTGTG TGCAGGGAGTTATGGCCGTGTCTAACTCTTGAGAGGCTGTGAGGATTCCGGAGTTCACACACC TCCGGCTTGGTCCCTGTACCTCACCTCCTTGGACT
6482	G	A	ATCTCCCCCTATCTCAGGGAGACACCTCCTACTGTGCCAGCATTTGTGACTCTTCTTTCACCCCC CTGCCCTTGGGTCCCTGGCCCTGGGATTGTTTGGGTGGAGGAGGGGAGTGGCTGCTGGCAGAATGG GGTGGAGGGGGGAGCGAAGCAGAGGGGGCGGGGAGTGGCCGGCTTTGAATATCTGTTGACCCC AGTTTCTCTGCCCCAGCTTATGTCTCTTCCCTCCCTCCTCTTCAAGCGTTAACTCCTTCTCTAA CTCGGGGGGAGAACGGGGCCAGGCCGCCAGGGGCA [G, A] GAGCTTTAGAATCAGGGTGACCCCCACCCCTACTCCCCAAGCACAGTACGGGCACACATACAAATG TGATGGTTTATCATTTGTATCTTTTGTGTTTGAAGGTGGGGGTCTAGGAGTCCAGAGGAGTGATG GGGTGCTGGAGGCTTCATTGGCAGCCTCCTGCCCTGAGTCTGGCTGGGGAGTCCCAGTTTTCTTAA GACTTGAATCCTGCCAGCAGTGGTGAGGCTGGGAGAGGCTCTTAGGAGGGACGGTGAGGCAGGGTG GAGCTTGGTACTAAGGATGGCGACCTAGGTCTCTAA
8066	C	G	TCTGTGATGGGTGCGTGACACCTGTCTTGGTGAACCTCACATCTTCTGCCTTGCTCCTGAGTGCA TGTGTGTGTTGCGCTCCATTTCTCTGGCCAGCCCGTGTATCTGCCTCCTGGCCTCTTCGGGCTTGT CTTCTTTTCTGTGTTCTGAGTTCAGGGGTGTGGGTTCCAGATCCCTGGCTGTTGCCAGTTAGCC CCATGTCTTCTATTTCTGACTCACCAGCAGCCCTGAGGTCTTTTCCCTGGAAGGGAGGAGTCAGG TGTGTGCTGTGGGTGGGGGAAGACTCCTGCCATC [C, G] TGCAGTGTGAGGCAGGTAAGTGGGATTCTCCTGAGGAGGATCCTTTTAGGTGAATCATTCTCCCCA GCTTTTCTGGCCTGCTCAGGTAGGCGATGGGCAAACGCTTGGGGGAGCAGCTGGCCTGGCCCTCC TCCCTTAGACTGAGACCGTAGCCAGGCACTGCTCCCACTGTGGGTGTGGACAACCTGACTCCCTCC CCTCCATACCCAGGGCTTCTTCCGCCGAGCATCCAGAAGAACATGGTGTACACGTGTACCCGGGA CAAGAACTGCATCATCAACAAGGTGACCCGGAACCG
8699	T	C	AAGTGGGCATGTCCAAGGAGTGTGAGTGCCATAGGGCAGGGGCCGAGTCCCGCCTCAGTTGGGGTC TCAGATGCTCCTAAAGACCAAGGGAGCAGGGCTCTGTGGATGTTTGTGCACATGCATGAACACGCA TGCCGTGGTGTGCGGGCTCACGGTTGAGGATGGTTTGTGTGTAGCTGCAAGGACCTGTTTGCAGT CTGGCTGGCTGTGTGTCCACGGGCAGTCTGTGCTCCGGGACCGTGTATGTGTAACCATTCCTGTT TCTGCACGTCTGGCTGTGTGTGCTTGCCTATGTGTG [T, C] GTGTGTGCATGCTCCAGGATGGCTTTCTTCCAGGCCGTGCTTGGTTTTGGGGTGGGGCTCAGAGGC ATAGGCAGTCCCTTCTGATTGTGAGTCTTAGGGGAGGGGCTTGAATTCAGAGGGGTGCTTGGCTGG ACTTATGTGTGTATGGGGGGGTGGAAGGGCTGGCACAAGGATCCAAAAGCCATTGTCTAGTTAAGC CTGGGATTTCAGAGTTGGAAGAAAGAAATTTGGGACTTCTCAGATCCCAGAGGAAACGGGGTTTCCACT TTGGGCTCAGCTGAGGCCTGATGGAGGGAGGGAGGG
12897	C	T	CAGTAGACGCATGTGGCTGGTGGCTGAGGTATGGATGGTGCAGACGTAGAACCTTTCCATCATTGT AGAAAATTCTATCAGACAGCATTGCTCCGGCCACCTGCCAGGTGGTCCCTCCGGGAGTGCTGGTGCG GAGTGCTGGTGCCGAGTGCTCAGAGTGGGTTGGGTTTCAGTCCCTGAACCCAGCATCCTCTGCAC CCAGATCCTGCGGATCTGCACGCGGTACACGCCGAGCAGGACCATGACCTTCTCGGACGGGCT GACCTGAACCGGACCCAGATGCACACGCTGGCTT [C, T] GGCCCCCTCACCGACCTGGTCTTTCGCTTCGCCAACCAGCTGCTGCCCTGGAGATGGATGATGCG GAGACGGGGCTGCTCAGCGCCATCTGCCTCATCTGCGGAGGTGGGCAGGGGGCTGGGTCTGGGGG CTGGGCTGGGACGGGGGTGCAGCCCTGGAGTCTCTTCCAGGGAGCTCTTTCAGGCCACCTCTGTTA GGTATCTCTAGAGGGCAGGGTCTGGTCTGCAACTACACAGCAAGGGGGCCATGTGGGGCTGGACT CCTGTTCCCGATTTCTGGGCAACACCCCTTCTAGGG
14442	C	T	TGTCTTGTGTGGGTCACTCAACAGCTCCTAGCTACAGTTTCCCTTCCGAGGGCGGGGATAACATT CGTGTTTACAGAGGGGTGCGGATGATCCCTAGCACACAGCACAGGGGAAGGAAGGGCTTGGCGTCT AGCCCAGGCCGGCAGTCTGGCCCTGGAGCCGAGTTCGGGACCACTTTGCCCATTTGCCACAGCC TCTGGACCTGGGGGCTTAAGAGAGCTGGCTCGTGTCAAAGAACTGAATCCCAAGAAAGATGCTAAT ATCAGCAGTATTGATCTTCCACCTCGAGCCAGGCT [C, T] GCTGGGGCTGGGGGTGGGAGGGCTGGCCAGCGTGTGACCTCTGCCCCCTCCTTCTCTGCAGGGG CTGAGCGGGTGATCACGCTGAAGATGGAGATCCCGGGCTCCATGCCGCTCTCATCCAGGAAATGT TGGAGAACTCAGAGGGCTGGACACTCTGAGCGGACAGCCGGGGGTGGGGGGGCG

09691220 - 101900

POSITION	Allele 1	Allele 2		Protein Position		
4084	C	G	Intron			
6482	G	A	Intron			
8066	C	G	Intron			
8699	T	C	Intron			
12897	C	T	Exon	237	F	L
14442	C	T	Intron			

006T0T = 022T6960